

**EDIT**  
Platform for  
Cybertaxonomy



## TaxEditor

User manual, appendix:

Platform output model for botany  
("CDM-light" 2.04)

Version: 14 Dezember 2019



## Contents

1	Introduction.....	2
	Current limitations:.....	2
	Implementation.....	2
	Conventions.....	2
2	Table definitions (export).....	3
	CommonNameFact.csv.....	3
	FactSources.csv.....	3
	GeographicAreaFact.csv.....	3
	HomotypicGroup.csv.....	3
	MediaFact.....	4
	Metadata.....	4
	NameFact.csv.....	4
	NameIdentifier.csv.....	4
	NameRelationship.....	4
	PersonOrTeam.csv.....	5
	PersonTeamRelation.csv.....	5
	Reference.csv.....	5
	ScientificName.csv.....	6
	SimpleFact.csv.....	8
	Specimen.csv.....	8
	SpecimenFact.....	9
	Synonym.....	9
	Taxon.....	9
	TaxonInteractionFact.....	10
	Additional tables (for processing in MS Access and Word).....	10
	Facts2Use (designates fact categories that should be processed and their sequence in the output).....	10
3	Schema.ini file.....	11
4	Processing CDM-Light with Microsoft Office programs.....	14
	4.1. Role of schema.ini.....	14
	4.2. Producing Print-Output (Word, PDF, HTML).....	14
	4.3. Producing data for World Flora Online.....	14



# 1 Introduction

## Current limitations:

- Single classification
- No concept relationships except partial and pro-parte synonyms and misapplied names (as flags in synonym.csv table)
- Note: All taxon nodes must have a name Rank assigned
- Hybrid formulas will be converted to empty name strings

## Implementation

- Export to csv tables
- Character encoding Unicode UTF-8 (Windows Characterset 65001)
- Choice if csv tables are packed into .zip file

## Conventions

- All IDs are UUIDs as designated in the CDM
- Sequence of attributes in tables must be maintained for usage in MS-Access etc.

## 2 Table definitions (export)

### CommonNameFact.csv

1	Fact_ID	UUID of Fact	507fcd90-0bcb-436a-8548-8de58e1f5b70	Text
2	Taxon_Fk	Pointer to Taxon_ID in Taxon.csv	adcc91d1-35e0-4516-a8b0-9760da4aeeb8	Text
3	FactText	The actual common name	Guayabilla	Memo
4	Language	Language designation taken from a term list	Spanish	Text
5	AreaLabel	An area designation taken from a term list	Colombia, Republic of	Text

### CondensedDistributionFact.csv

1	Fact_ID	UUID of Fact		Text
2	Taxon_Fk	Pointer to Taxon_ID in Taxon.csv		Text
3	FactText	Condensed distribution		Memo

### FactSources.csv

1	Fact_Fk	Pointer to Fact_ID *Fact.csv	13d596d7-ec0f-4d5c-a4d6-da6e4d718442	Text
2	Reference_Fk	Pointer to Reference_ID in Reference.csv	b7197851-6bcc-4861-9db6-0a502837d072	Text
3	NameInSource_Fk	The scientific name used in the source fact	812097ff-6dfc-4061-9be6-7e384c31b0f4	Memo
4	FactType	may be SimpleFact, GeographicAreaFact, SpecimenFact or CondensedDistributionFact	SimpleFact	Text

### GeographicAreaFact.csv

1	Fact_ID	UUID of Fact	44d8e6e0-73f7-4b42-b722-7f2aeb45d44e	Text
2	Taxon_Fk	Taxon Pointer	3b1e2257-da35-44ea-87d5-2e7e0e55a017	Text
3	AreaLabel	The name of the geographic area	Ecuador	Text
4	StatusLabel	Occurrence status within the geographic area	native	Text

### HomotypicGroup.csv

1	HomotypicGroup_ID	UUID of homotypic Group	15360909-e644-4990-8dd2-b03f1f3c79e8	Text
---	-------------------	-------------------------	--------------------------------------	------

2	HomotypicGroupString	Concatenated output string for homotypic group (names and citations) + status + some name relations (e.g. “non”)	Echinocactus simpsonii Engelm.; Pediocactus simpsonii (Engelm.) Britton & Rose; Mammillaria simpsonii (Engelm.) M.E. Jones; Pediocactus simpsonii var. simpsonii	Memo
3	TypeSpecimenString	Concatenated output of all SpecimenCitations in the ScientificName table for the names belonging to the homotypic group		
4	TypeStatementsString	Concatenated output of all (normally: one) TypeStatements in the ScientificName table for the names belonging to the homotypic group		Memo

### Identifier.csv

1	ForeignKey	UUID of identified object in the referenced table	b3dd1aed-92e0-4058-be8e-4a3956d9f621	Text
2	ReferencedTable	The table holding the record of the identified object	ScientificName	
3	ExternalIdentifier	The actual identifier in the external system	wfo-2142232	Text
4	IdentifierType	The respective external identifier system	WFO Name Identifier	Text

### MediaFact

This is a table for facts containing media and no text

1	Fact_ID			Text
2	Taxon_Fk			Text
3	Name_Fk			
4	MediaURI			

### Metadata

1	EditInstance_ID	Identifier for the CDM Instance (database)		Text
2	EditInstanceName	Name of the CDM instance (database)		Text

### NameFact.csv

(Protologue-links are covered in the ScientificName table)



1	Fact_ID			Text
2	Name_Fk			Text
3	FactText			Memo
4	Language			Text
5	MediaURI			Text
6	FactCategory			Text

**NameRelationship**

1	Name1_Fk	UUID of first scientific name	666013f3-dd63-45ae-a118-412b7c113816	Text
2	Name2_Fk	UUID of second scientific name	25440686-bb1a-44e8-a5b6-8c0a67d84c10	Text
3	NameRelationshipType	description of relationship	is basionym for	Text

**PersonOrTeam.csv**

1	PersonOrTeam_ID	The UUID of the Author or Team	81058813-47f8-42ab-905f-de14b32291ce	Text
2	AbbrevNames	Nomenclatural abbreviation of author or team	T.Moore & Mast. [empty for bibliography-only authors]	Text
3	FullNames	Bibliographic citation of author(s)	Moore, T. & Masters, M.T.	Text
4	PersonOtherNames	for single person only		Text
5	PersonFamilyNames	for single person only		Text
6	PersonPrefix	for single person only		Text
7	PersonSuffix	for single person only		Text

**PersonTeamRelation.csv**

1	AuthorTeam_Fk	UUID of a team in PersonOrTeam.csv	ab34e611-6e60-4dbb-b6d9-a53ce06e8dde	Text
2	Author_Fk	UUID of an individual person in PersonOrTeam.csv	77181551-8f22-4a59-8bdb-5c721c55668e	Text
3	SequenceNumber	Position in the team	1	Short

**Reference.csv**

1	Reference_ID	The UUID for the nomenclatural or bibliographic reference	189bde42-f464-4a55-b773-365886751d77	Text
2	BibliographicShortCitation	In-text reference	Allen (2014) Allen & Seymore (2014) Allen & al. (2014)	Text



3	Title			Memo
4	AbbrevTitle			Text
5	DatePublished			Text
6	Edition			Text
7	Editor			Text
8	ISBN			Text
9	ISSN			Text
10	Organisation			Text
11	Pages			Text
12	PlacePublished			Text
13	Publisher			Text
14	ReferenceAbstract			Memo
15	SeriesPart			Text
16	Volume			Text
17	Year			Text
18	FullAuthor	= FullNames in table PersonOrTeam		Text
19	Author_Fk	UUID of person in PersonOrTeam.csv		Text
20	InReference			Text
21	Institution			Text
22	LSID			Text
23	School	Attribute in use for publications of type Thesis		Text
24	ReferenceType			Text
25	URI	URI for the full publication if available on-line		Text

**ScientificName.csv**

1	Name_ID	UUID for the Name in the CDM Instance		Text
2	LSID			Text
3	Rank	VerbatimRank (Code – Latin)	Species	Text
4	RankSequence	Numeric Rank Identifier (ordered) using the cdm rank class see <a href="http://cybertaxonomy.eu/cdmlib/apidocs/eu/etaxonomy/cdm/model/name/RankClass.html">http://cybertaxonomy.eu/cdmlib/apidocs/eu/etaxonomy/cdm/model/name/RankClass.html</a>	44	Short
5	FullNameWithAuthors	For Tropicos name checking – Authors in Tropicos	Coccoloba nicaraguensis Standl. & L.O. Williams	Text



		Standard (no blanks between initials, but a blank between lastname and initials, and some other exceptions)		
6	FullNameNoAuthors	For Tropicos name checking (= nameCache)		Text
7	GenusOrUninomial	Genus or monomial Name		Text
8	InfragenericRank	Abbreviated rank term for subdivision of genus	z.B. "subg."	Text
9	InfraGenericEpithet			Text
10	SpecificEpithet			Text
11	InfraspecificRank	Abbreviated rank term for infraspecies	z.B. "var."	Text
12	InfraSpecificEpithet	Infraspecific epithet		Text
13	BasionymExAuthorTeam_Fk			Text
14	BasionymAuthorTeam_Fk	Pointer to basionym author(s)		Text
15	PublishingExAuthorTeam_Fk			Text
16	PublishingAuthorTeam_Fk	Pointer to combination author(s)		Text
17	AuthorTeamString	Author team string including basionym authors and ex authors		Text
18	Reference_Fk			Text
19	PublicationType	Article / Book / Book-section etc.	In CDM it is no problem to provide all reference types	Text
20	AbbreviatedTitle	BPH- or TL2-Type abbreviation (A/B) or	abbrev title of the "higher" (see below) nomenclatural reference, NOT the cache!	Text
21	FullTitle	Bibliographic citation of Book (for B-s)	See above, not abbreviated	Memo
22	AbbreviatedInRefAuthor	Abbreviated author of the "higher" reference	Higher references are Books, Journals, etc., NOT (Book)Section, Article, but careful with PrintSeries which are usually not higher references in terms of citations	Text
23	FullRefAuthor	same as AbbreviatedRefAuthor but not abbreviated		Text
24	Collation	Volume (issue): Page - OtherPage, Figure (Standartzitat)		Text
25	Volumelssue	In cdm volume and issue are in one field		Text
26	Detail	Incl. page, other page, figure, ... (a.k.a. nomenclaturalMicroReference)		Text
27	DatePublished			Text
28	YearPublished			Text
29	VerbatimDate			Text





30	ProtologueURI	Concatenated list of URLs given for the protologue or nomenclatural act.		Text
31	NomenclaturalStatus	Concatenated list of status' assigned to the name		Text
32	NomenclaturalStatusAbbreviation			Text
33	HomotypicGroup_Fk	Can be used as an arbitrary unique identifier for the homotypic group, or as a pointer to the concatenated homotypic group string		Text
34	HomotypicGroupSequenceNumber			Short
35	ProtologueTypeStatement	The original (verbatim) text establishing the type of this name in the protologue		Text
36	TypeSpecimens	Text composed from individual type specimens assigned to this name		Text
37	TypeStatements	Free text. May include citation of the type(s) of this name, and may describe actions taken in the typification process		Text

**SimpleFact.csv**

1	Fact_ID			Text
2	Taxon_Fk	Taxon Pointer		Text
3	FactText			Memo
4	Language	For every representation of a Multilanguage text create a new fact		Text
5	MediaURI	URI; URI; ...		Text
6	FactsCategory			Text

**Specimen.csv**

1	Specimen_ID			Text
2	SpecimenCitation	Locality, Collector, Number, Date, Herbarium code and herbarium number, if present.	titleCache	Memo
3	FieldUnitCitation	Specimen citation without herbarium code and number		
4	Locality			Memo
5	Country			Text



6	AreaCategory1	Usually abbreviated		Text
7	AreaName1			Text
8	AreaCategory2	Usually abbreviated		Text
9	AreaName2			Text
10	AreaCategory3	Usually abbreviated		Text
11	AreaName3			Text
12	FurtherAreas			Text
13	CollectorString		Should be the collectorString if it is implemented (?) otherwise titleCache	Text
14	CollectorNumber			Text
15	CollectionDate			Text
16	SpecimenImageURIs	URI; URI; ...		Memo
17	HerbariumAbbrev			Text
18	MediaSpecimenURL	Specimen only available as media?		Text
19	PreferredStableID	CETAF GUID		Memo
20	Barcode			Text
21	CatalogueNumber			Text
22	AccessionNumber			Text

**SpecimenFact**

1	Fact_ID			Text
2	Taxon_Fk	Taxon Pointer		Text
3	Specimen_Fk			Text
4	SpecimenDescription			Memo
5	SpecimenNotes			Memo

**Synonym**

1	Synonym_ID			Text
2	Taxon_Fk	Taxon Pointer		Text
3	Name_Fk			Text
4	SynSecReference_Fk			Text
5	SynSecReference	Bibliographic short citation of syn-sec.-reference		Memo
6	SecReference_Fk			Text

7	SecReference			Memo
8	IsProParteSynonym			Bit
9	IsPartial			Bit
10	isMisapplied			Bit
11	Published	Marker for published names		Bit

**Taxon**

1	Taxon_ID	Taxon Identifier		Text
2	Classification_ID			Text
3	ClassificationName			Text
4	Name_Fk			Text
5	Parent_Fk	Uuid of the parent taxon		Text
6	SecReference_Fk			Text
7	SecReference	Bibliographic short citation of sec.-reference		Memo
8	SortIndex	Sortorder determined at time of export		Short
9	Excluded	Marker for excluded taxa		Bit
10	ExcludedNotes	Reason for the exclusion of the taxon		Memo
11	Published	Marker for published taxa		Bit
12	Unplaced	Marker for unplaced taxa		Bit

**TaxonInteractionFact**

1	Fact_ID			Text
2	Taxon_Fk	Taxon Pointer		Text
3	Taxon2_Fk	Taxon Pointer		Text
4	Description			Memo

Additional tables (for processing in MS Access and Word)

**Facts2Use (designates fact categories that should be processed and their sequence in the output)**

1	FactCategory	Name of the fact category		Text
---	--------------	---------------------------	--	------



2	FactSequence	Position of the facts belonging to this categorie in the output		Short
3	FactPrompt	Category name in output		Text



### 3 Schema.ini file

[CommonNameFact.csv] Format=CSV Delimited CharacterSet=65001 Col1=Fact_ID Text Col2=Taxon_FK Text Col3=FactText Memo Col4=Language Text Col5=AreaLabel Text	Col3=Name nSource_FK Text Col4=FactType Text	[Metadata.csv] Col1=EditInstance_ID Text Col2=EditInstanceName Text [NameFact.csv] Format=CSV Delimited CharacterSet=65001 Col1=Fact_ID Text Col2=Name_FK Text Col3=FactText Memo Col4=Language Text Col5=MediaURI Text Col6=FactCategory Text	[PersonOrTeam.csv] Format=CSV Delimited CharacterSet=65001 Col1=PersonOrTeam_ID Text Col2=AbbrevNames Text Col3=FullNames Text Col4=PersonOtherNames Text Col5=PersonFamilyName Text Col6=PersonPrefix Text Col7=PersonSuffix Text	Col6=Edition Text Col7=Editor Text Col8=ISBN Text Col9=ISSN Text Col10=Organisation Text Col11=Pages Text Col12=Place Published Text Col13=Publisher Text Col14=Reference Abstract Memo Col15=Series Part Text Col16=Volume Text Col17=Year Text Col18=FullAuthor Text Col19=Author_Fk Text Col20=InReference Text Col21=Institution Text Col22=LSID Text Col23=School Text Col24=Reference Type Text Col25=URI Text	Col4=RankSequence Short Col5=FullName WithAuthors Text Col6=FullName NoAuthors Text Col7=GenusOrUninomial Text Col8=InfrgenericRank Text Col9=InfraGenericEpithet Text Col10=SpecificEpithet Text Col11=InfrspecificRank Text Col12=InfraSpecificEpithet Text Col13=BasionymExAuthor Team_Fk Text Col14=BasionymAuthor Team_Fk Text Col15=PublishingExAuthor Team_Fk Text Col16=PublishingAuthor Team_Fk Text Col17=Author TeamString Text Col18=Reference_Fk Text Col19=Publication Type Text Col20=Abbreviated Title Text Col21=FullTitle Memo Col22=AbbreviatedIn Reference Author Text
[CondensedDistributionFact.csv] Format=CSV Delimited CharacterSet=65001 Col1=Fact_ID Text Col2=Taxon_FK Text Col3=FactText Memo	Col1=Fact_ID Text Col2=Taxon_FK Text Col3=AreaLabel Text Col4=StatusLabel Text	[Identifier.csv] Format=CSV Delimited CharacterSet=65001 Col1=Foreign Key Text Col2=Referenced Table Text Col3=External Identifier Text Col4=Identifier Type Text	[PersonTeamRelation.csv] Format=CSV Delimited CharacterSet=65001 Col1=Author Team_Fk Text Col2=Author_Fk Text Col3=Sequence Number Short	Col18=FullAuthor Text Col19=Author_Fk Text Col20=InReference Text Col21=Institution Text Col22=LSID Text Col23=School Text Col24=Reference Type Text Col25=URI Text	Col13=BasionymExAuthor Team_Fk Text Col14=BasionymAuthor Team_Fk Text Col15=PublishingExAuthor Team_Fk Text Col16=PublishingAuthor Team_Fk Text Col17=Author TeamString Text Col18=Reference_Fk Text Col19=Publication Type Text Col20=Abbreviated Title Text Col21=FullTitle Memo Col22=AbbreviatedIn Reference Author Text
[Facts2use.csv] Format=CSV Delimited CharacterSet=65001 Col1=FactCategory Text Col2=FactSequence Short Col3=FactPrompt Text	[HomotypicGroup.csv] Format=CSV Delimited CharacterSet=65001 Col1=Homotypic Group_ID Text Col2=Homotypic GroupString Memo Col3=TypeSpecimen String Memo Col4=TypeStatements String Memo	[NameRelationship.csv] Format=CSV Delimited CharacterSet=65001 Col1=Name1_Fk Text Col2=Name2_Fk Text Col3=NameRelationship Type Text	[Reference.csv] Format=CSV Delimited CharacterSet=65001 Col1=Reference_ID Text Col2=Bibliographic ShortCitation Text Col3=Title Memo Col4=AbbrevTitle Text Col5=Date Published Text	[ScientificName.csv] Format=CSV Delimited CharacterSet=65001 Col1=Name_ID Text Col2=LSID Text Col3=Rank Text	Col13=BasionymExAuthor Team_Fk Text Col14=BasionymAuthor Team_Fk Text Col15=PublishingExAuthor Team_Fk Text Col16=PublishingAuthor Team_Fk Text Col17=Author TeamString Text Col18=Reference_Fk Text Col19=Publication Type Text Col20=Abbreviated Title Text Col21=FullTitle Memo Col22=AbbreviatedIn Reference Author Text
[FactSources.csv] Format=CSV Delimited CharacterSet=65001 Col1=Fact_Fk Text Col2=Reference_Fk Text	[MediaFact.csv] Format=CSV Delimited CharacterSet=65001 Col1=Fact_ID Text Col2=Taxon_FK Text Col3=Name_Fk Text Col4=MediaURI Text				



Col23=FullInRefAuthor Text	Col5=Media URI Text	Col19=PreferredStableID Text	Col8=IsProParteSynonym Bit	Col4=Description Memo
Col24=Collation Text	Col6=FactCategory Text	Col20=Barcode Text	Col9=IsPartial Bit	
Col25=VolumeIssue Text	[Specimen.csv]	Col21=CatalogueNumber Text	Col10=IsMisapplied Bit	
Col26=Detail Text	Format=CSV Delimited	Col22=AccessionNumber Text	Col11=Published Bit	
Col27=DatePublished Text	CharacterSet=65001		Taxon.csv]	
Col28=YearPublished Text	Col1=Specimen_ID Text	[SpecimenFact.csv]	Format=CSV Delimited	
Col29=VerbatimDate Text	Col2=SpecimenCitation Memo	Format=CSV Delimited	CharacterSet=65001	
Col30=ProtocolURI Text	Col3=FieldUnitCitation Memo	CharacterSet=65001	Col1=Taxon_ID Text	
Col31=NomenclaturalStatus Text	Col4=Locality Memo	Col1=Fact_ID Text	Col2=Classification_ID Text	
Col32=NomenclaturalStatusAbbreviation Text	Col5=Country Text	Col2=Taxon_Fk Text	Col3=ClassificationName Text	
Col33=HomotypicGroup_Fk Text	Col6=AreaCategory1 Text	Col3=Specimen_Fk Text	Col4=Name_Fk Text	
Col34=HomotypicGroupSequenceNumber Short	Col7=AreaName1 Text	Col4=SpecimenDescription Memo	Col5=Parent_Fk Text	
Col35=ProtocologueTypeStatement Text	Col8=AreaCategory2 Text	Col5=SpecimenNotes Memo	Col6=SecReference_Fk Text	
Col36=TypeSpecimens Memo	Col9=AreaName2 Text		Col7=SecReference Text	
Col37=TypeStatements Memo	Col10=AreaCategory3 Text	[Synonym.csv]	Col8=SortIndex Short	
[SimpleFact.csv]	Col11=AreaName3 Text	Format=CSV Delimited	Col9=Excluded bit	
Format=CSV Delimited	Col12=FurtherAreas Text	CharacterSet=65001	Col10=ExcludedNotes Memo	
CharacterSet=65001	Col13=CollectorString Text	Col1=Synonym_ID Text	Col11=Published bit	
Col1=Fact_ID Text	Col14=CollectorNumber Text	Col2=Taxon_Fk Text	Col12=Unplaced bit	
Col2=Taxon_Fk Text	Col15=CollectionDate Text	Col3=Name_Fk Text		
Col3=FactText Memo	Col16=SpecimenImageURIs Memo	Col4=SynSecReference_Fk Text	[TaxonInteractionFact.csv]	
Col4=Language Text	Col17=HerbariumAbbrev Text	Col5=SynSecReference Memo	Format=CSV Delimited	
	Col18=MediaSpecimenURI Text	Col6=SecReference_Fk Text	CharacterSet=65001	
		Col7=SecReference Memo	Col1=Fact_ID Text	
			Col2=Taxon_Fk Text	
			Col3=Taxon2_Fk Text	

## 4 Processing CDM-Light with Microsoft Office programs

### 4.1. Role of schema.ini

For import in MS-Access (and Excel ?) place schema.ini as a text-only file in the same directory with your access database and the CDM-light .cvs tables. This ensures that the character encoding and the data type of the fields in the tables are recognised (otherwise these programs try to detect the data type from the first few records).

Schema.ini defines attribute data types for import into MS-Access according to the Microsoft Jet data types: Bit, Byte, Short [= Integer], Long, Currency, Single, Double, DateTime, Text [≤ 255 characters], Memo [≤ 2 GB text BLOB] (see [https://en.wikibooks.org/wiki/JET\\_Database/Data\\_types](https://en.wikibooks.org/wiki/JET_Database/Data_types)). Default type is Text.

### 4.2. Producing Print-Output (Word, PDF, HTML)

To produce a textual checklist of the content: Manual-EDIT-Platform-Appendix-Checklist-Document-Procedure.docx (in preparation)

### 4.3. Producing data for World Flora Online

To produce a Darwin Core Archive format file to replace a family treatment in the World Flora Online taxonomic backbone: Manual-EDIT-Platform-Appendix-WFO-Backbone-Procedure.docx

To produce a Darwin Core Archive format file for Content to be added to the World Flora Online database: Manual-EDIT-Platform-Appendix-WFO-Content-Procedure.docx (in preparation)